

INFO 6210

Data Management and Database

Final Project 2 - NoSQL

I chose my domain as soccer. I scraped the dataset from Kaggle.
<https://www.kaggle.com/ankitasahni/exploring-soccer-analysis/data>.

In this project, I have converted my project 1 SQL database to NoSQL successfully. I connected tables to one table in SQL, then import it to csv file and put it into mongoDB.

Sample Data:

Player

index	Name	National_id	Height	Weight	Preferred_Foot	Birth_Date	Age	Preferred_Position	Work_Rate
0	Cristiano Ronaldo	a6	185 cm	80 kg	Right	02/05/1985	32	LW/ST	High / Low
1	Lionel Messi	a5	170 cm	72 kg	Left	06/24/1987	29	RW	Medium / Medium
2	Neymar	a3	174 cm	68 kg	Right	02/05/1992	25	LW	High / Medium
3	Luis Su rez	a11	182 cm	85 kg	Right	01/24/1987	30	ST	High / Medium
4	Manuel Neuer	a2	193 cm	92 kg	Right	03/27/1986	31	GK	Medium / Medium
5	De Gea	a1	193 cm	82 kg	Right	11/07/1990	26	GK	Medium / Medium
6	Robert Lewandowski	a12	185 cm	79 kg	Right	08/21/1988	28	ST	High / Medium
7	Gareth Bale	a24	183 cm	74 kg	Left	07/16/1989	27	RW	High / Medium
8	Zlatan Ibrahimovi	a21	195 cm	95 kg	Right	10/03/1981	35	ST	Medium / Low
9	Thibaut Courtois	a4	199 cm	91 kg	Left	05/11/1992	24	GK	Medium / Medium
10	Jrme Boateng	a2	192 cm	90 kg	Right	09/03/1988	28	CB	Medium / Medium

Nationality(Country)

index	National_id	Nationality_Name
0	a1	Spain
1	a2	Germany
2	a3	Brazil
3	a4	Belgium
4	a5	Argentina
5	a6	Portugal
6	a7	France
7	a8	Italy
8	a9	England
9	a10	Netherlands

Club

index	Club	National_id	Club_Position	Club_Kit	Club_Joining
0	Real Madrid	a6	LW	7	07/01/2009
1	FC Barcelona	a5	RW	10	07/01/2004
2	FC Barcelona	a3	LW	11	07/01/2013
3	FC Barcelona	a11	ST	9	07/11/2014
4	FC Bayern	a2	GK	1	07/01/2011
5	Manchester Utd	a1	GK	1	07/01/2011
6	FC Bayern	a12	ST	9	07/01/2014
7	Real Madrid	a24	RW	11	09/02/2013
8	Manchester Utd	a21	ST	9	07/01/2016
9	Chelsea	a4	GK	13	07/26/2011

Full data (the table that made up of “Player”, “Nationality” and “Club”)

index	"Name"	"National_id"	"Height"	"Weight"	"Preferred_Foot"	"Birth_Date"	"Age"	"Preferred_Position"	"Work_Rate"	"Aggression"	"Reactivi
0	Cristiano Ronaldo	a6	185 cm	80 kg	Right	02/05/1985	32	LW/ST	High / Low	63	
1	Lionel Messi	a5	170 cm	72 kg	Left	06/24/1987	29	RW	Medium / Medium	48	
2	Neymar	a3	174 cm	68 kg	Right	02/05/1992	25	LW	High / Medium	56	
3	Luis Su rez	a11	182 cm	85 kg	Right	01/24/1987	30	ST	High / Medium	78	
4	Manuel Neuer	a2	193 cm	92 kg	Right	03/27/1986	31	GK	Medium / Medium	29	
5	De Gea	a1	193 cm	82 kg	Right	11/07/1990	26	GK	Medium / Medium	38	
6	Robert Lewandowski	a12	185 cm	79 kg	Right	08/21/1988	28	ST	High / Medium	80	
7	Gareth Bale	a24	183 cm	74 kg	Left	07/16/1989	27	RW	High / Medium	65	
8	Zlatan Ibrahimovi	a21	195 cm	95 kg	Right	10/03/1981	35	ST	Medium / Low	84	
9	Thibaut Courtois	a4	199 cm	91 kg	Left	05/11/1992	24	GK	Medium / Medium	23	

Data from social media sites :

Tweets

id	created_at	tweets	user.id	user.location	retweet_count
1	2018-04-17 17:50:07	RT @sthabieeyr_77: Some people are talented #Lionel Mes...	834154403930238979	â– South Africa â– , Ladysmithâ	1
2	2018-04-17 03:42:41	RT @NoticiasVideos2: Atletico #Madrid #Vs #Barcelona #2...	936567941121888258	Miami, FL	2
4	2018-04-16 23:13:03	RT @NoticiasVideos2: Atletico #Madrid #Vs #Barcelona #2...	801715146590711808	<null>	2
5	2018-04-16 18:33:40	Atletico #Madrid #Vs #Barcelona #2017 1-2 #Messi #Vs ...	801715146590711808	<null>	2
6	2018-04-16 08:30:02	#Lionel Messi Vs #Cristiano Ronaldo https://t.co/xhWLIT...	984455546504990720	<null>	0
7	2018-04-16 05:06:14	Cristiano Ronaldo#CR7#The Beast#Legend#And The World...	862646148489773056	Islamabad, Pakistan	0
8	2018-04-15 16:10:56	#BabeKasihTau Membandingkan Reaksi Cristiano Ronaldo da...	2306171660	Indonesia	0
9	2018-04-15 14:41:13	#cristiano ronaldo#Lionel messi	974246516050325504	snap:nicobj63	0
10	2018-04-13 20:56:58	RT @LibanoCule: #Lionel #Messi goal against Leganes , f...	970896576792195073	<null>	3

Hashtags

	id	created_at	tag	player_name
1	2018-04-17 17:50:07	Lionel	Lionel Messi	
2	2018-04-17 03:42:41	Madrid	Lionel Messi	
2	2018-04-17 03:42:41	Vs	Lionel Messi	
2	2018-04-17 03:42:41	Barcelona	Lionel Messi	
2	2018-04-17 03:42:41	Messi	Lionel Messi	
2	2018-04-17 03:42:41	Vs	Lionel Messi	
2	2018-04-17 03:42:41	Griezmann	Lionel Messi	
2	2018-04-17 03:42:41	All	Lionel Messi	
4	2018-04-16 23:13:03	Madrid	Lionel Messi	
4	2018-04-16 23:13:03	Vs	Lionel Messi	

Tweets full data (the table made up of “tweets” and “hashtags”)

	id	created_at	tweets	user_id	user_location	tag
2642	2018-04-17 17:45:10	RT @NeymarJrSite: NÃ£o tire o olho do @neymarjr ...@Que...	976187699857108994	Bloemfontein, South Africa	stopmotion	
2292	2018-04-17 17:45:10	RT @NeymarJrSite: NÃ£o tire o olho do @neymarjr ...@Que...	976187699857108994	Bloemfontein, South Africa	stopmotion	
2292	2018-04-17 17:45:10	RT @NeymarJrSite: NÃ£o tire o olho do @neymarjr ...@Que...	976187699857108994	Bloemfontein, South Africa	neymar	
1942	2018-04-17 17:45:10	RT @NeymarJrSite: NÃ£o tire o olho do @neymarjr ...@Que...	976187699857108994	Bloemfontein, South Africa	neymar	
1592	2018-04-17 17:45:10	RT @NeymarJrSite: NÃ£o tire o olho do @neymarjr ...@Que...	976187699857108994	Bloemfontein, South Africa	neymar	
192	2018-04-17 17:45:10	RT @NeymarJrSite: NÃ£o tire o olho do @neymarjr ...@Que...	976187699857108994	Bloemfontein, South Africa	neymar	
542	2018-04-17 17:45:10	RT @NeymarJrSite: NÃ£o tire o olho do @neymarjr ...@Que...	976187699857108994	Bloemfontein, South Africa	neymar	
1242	2018-04-17 17:45:10	RT @NeymarJrSite: NÃ£o tire o olho do @neymarjr ...@Que...	976187699857108994	Bloemfontein, South Africa	neymar	
2642	2018-04-17 17:45:10	RT @NeymarJrSite: NÃ£o tire o olho do @neymarjr ...@Que...	976187699857108994	Bloemfontein, South Africa	neymar	
892	2018-04-17 17:45:10	RT @NeymarJrSite: NÃ£o tire o olho do @neymarjr ...@Que...	976187699857108994	Bloemfontein, South Africa	stopmotion	

Collections in mongoDB:

Soccer full data

The screenshot shows the MongoDB Compass interface with the following details:

- My Cluster**: Localhost:27017 STANDALONE
- local.soccer_fulldata**: 14.1k DOCUMENTS, TOTAL SIZE 5.4MB, AVG. SIZE 406B, 1 INDEXES, TOTAL SIZE 136.0KB, AVG. SIZE 136.0KB.
- Documents** tab selected.
- Filter**: { field: 'value' }
- Fields** (expanded):
 - _id**: ObjectId("5ad4f9e50c9d75420d66dac6")
 - index: 0**
 - Name**: "Cristiano Ronaldo"
 - National_id**: "a6"
 - Height**: "185 cm"
 - Weight**: "80 kg"
 - Preferred_Foot**: "Right"
 - Birth_Date**: "02/05/1985"
 - Age**: 32
 - Preferred_Position**: "LW/ST"
 - Work_Rate**: "High / Low"
 - Aggression**: 63
 - Reactions**: 96
 - Vision**: 85
 - Crossing**: 84
 - Speed**: 92
 - Nationality_Name**: "Portugal"
 - Club**: "Real Madrid"
 - Club_Position**: "LW"
 - Club_Joining**: "07/01/2009"
- OPTIONS** and **FIND** buttons.

Soccer twitter

	id Int32	created_at String	tweets String	user_id Mixed	user_location Mixed
1	238	"2018-04-26 14:14:18"	"TIKI TAKA #nba #nhl #mlb #n	26035191	"Brazil"
2	284	"2018-04-21 01:30:39"	"PERFIL PSICOLÀGICO DE #GARI	1129277120	null
3	131	"2018-04-25 20:08:41"	"RT @FootballAddict: Marco #,	613371640	"USA, Florida."
4	202	"2018-04-26 17:40:25"	"RT @XtraminuteTV: Even with	986291031632240640	null
5	8	"2018-04-26 13:17:07"	"Lionel Messi puÃ³ registra	717039627916484608	"Italia"
6	77	"2018-04-26 16:41:48"	"RT @myfootballtales: Follow	975786316901298176	"Roma, Lazio"
7	217	"2018-04-26 16:39:55"	"RT @FantoAuction: RM v PSG	987712754680246274	"Sebokeng, South Afric
8	60	"2018-04-16 14:04:07"	"LÃ±deres goleadores de la L	876796796768485377	"Cuba"
9	3	"2018-04-26 19:11:15"	"#Lionel AndrÃ¡es #Messi Cuc	1101320424	"KÃ¶ln I Frankfurt"
10	292	"2018-04-22 04:24:02"	"Robert Lewandowski becomes	4097611282	"Deutschland"
11	274	"2018-04-23 04:42:10"	"#spain #spanish #harryfan #	3253704684	"Pakistan"
12	228	"2018-04-26 15:05:11"	"#NeymarÃ¢s move from #Bar-	824169988626018304	"United Kingdom"
13	144	"2018-04-25 19:44:18"	"RT @FootballAddict: Marcelo	714359486278795264	null
14	41	"2018-04-21 20:06:41"	"Lionel Messi Goal Vs Sevill	815451895691087874	"Houston, TX"

Code:

Loading data to mongoDB

```
In [1]: import pymongo
import pprint

In [2]: import pandas as pd
from pymongo import MongoClient
import json

def mongoimport(csv_path, db_name, coll_name, db_url='localhost', db_port=27017):
    client = MongoClient(db_url, db_port)
    db = client[db_name]
    coll = db[coll_name]
    data = pd.read_csv(csv_path, encoding='latin1')
    payload = json.loads(data.to_json(orient='records'))
    coll.remove()
    coll.insert(payload)
    return coll.count()
```

Social media data

1. Aggregate social media data

```
In [4]: import psycopg2
conn = psycopg2.connect(database='postgres', user='postgres', password='')
df_k=pd.read_sql('''select id,created_at,tweets,user_id,user_location,retweet_count,"Urls","media",player_name,array_agg(tweet_text) as tweet_text from full_tweet_data group by id,created_at,tweets,user_id,user_location,retweet_count,"Urls","media",player_name;''', con=conn)
```

In [5]:	df_k
	0 238 26 14:14:18 #nfl #ufc #surf #sk8 ... 26035191 Brazil 0 'https://t.co/7Jdt2QlaA0', 'expanded_... None Neymar
1 284	2018-04-21 01:30:39 PERFIL PSICOLÓGICO DE #GARETH #BALE #PSICOL... 1129277120 None 0 'https://t.co/LwXytB8GsM', 'expanded_... [{"id": 987503848477609987, "id_str": "987503848477609987"}] Gareth Bale
2 131	2018-04-25 20:08:41 RT @FootballAdictt: Marco #Asensio Goal HD - B... 613371640 USA, Florida. 8 [] None Cristian Ronaldo
3 202	2018-04-26 17:40:25 RT @XtraminuteTV: Even with a broken foot, Ney... 986291031632240640 None 0 [] [{"id": 986814607254872064, "id_str": "986814607254872064"}] Neymar
4 8	2018-04-26 13:17:07 Lionel Messi puÃ² registrare marchio per abbigliamento 717039627916484608 Italia 0 'https://t.co/lCH9bnnyfKX', 'expanded_... None Lionel Messi
5 77	2018-04-26 16:41:48 RT @myfootballtales: Follow us on instagram as... 975786316901298176 Roma, Lazio 1 [] None Cristiano Ronaldo
6 217	2018-04-26 RT @FantoAuction: RM v DSG - goals goals con... 987712754680246274 Sebokeng, South Africa 1 [] None Neymar

Soccer data

1. Use JOINS to denormalize tables

```
In [18]: import psycopg2
conn = psycopg2.connect(database='postgres', user='postgres', password='')
df1=pd.read_sql('''
SELECT A.* , B."Nationality_Name", C."Club",C."Club_Position", C."Club_Joining"
FROM
(
    (SELECT * FROM soccer_player) A
    FULL OUTER JOIN
    (SELECT * FROM soccer_nationality) B on A."National_id" = B."National_id"
    FULL OUTER JOIN
    (SELECT * FROM soccer_club) C on A."index" = C."index"
);''', con=conn)
```

In [19]:	df1													
Out[19]:														
	level_0	index	Name	National_id	Height	Weight	Prefferred_Foot	Birth_Date	Age	Prefferred_Position	Work_Rate	Aggression	Reactions	Vision
0	0.0	0.0	Cristiano Ronaldo	a6	185 cm	80 kg	Right	02/05/1985	32.0	LW/ST	High / Low	63.0	96.0	85.
1	1.0	1.0	Lionel Messi	a5	170 cm	72 kg	Left	06/24/1987	29.0	RW	Medium / Medium	48.0	95.0	90.
2	2.0	2.0	Neymar	a3	174 cm	68 kg	Right	02/05/1992	25.0	LW	High / Medium	56.0	88.0	80.
3	3.0	3.0	Luis Suárez	a11	182 cm	85 kg	Right	01/24/1987	30.0	ST	High / Medium	78.0	93.0	84.
4	4.0	4.0	Manuel Neuer	a2	193 cm	92 kg	Right	03/27/1986	31.0	GK	Medium / Medium	29.0	85.0	70.
5	5.0	5.0	De Gea	a1	193 cm	82 kg	Right	11/07/1990	26.0	GK	Medium / Medium	38.0	88.0	68.

2.Export to CSV

```
In [26]: df1.to_csv('FullData1.csv', encoding='utf-8', index=False)

In [41]: myfile1 = pd.read_csv('/Users/yizheliu/Desktop/DataBase/FullData1.csv', encoding='latin1')

In [42]: myfile1.duplicated().sum()

Dut[42]: 0

In [43]: myfile1.isnull().sum()

Dut[43]: index          0
Name            0
National_id    0
Height          0
Weight          0
Preferred_Foot 0
Birth_Date      0
Age             0
Preferred_Position 0
Work_Rate        0
Aggression       0
Reactions         0
Vision           0
Crossing          0
Speed             0
Nationality_Name 0
Club             0
Club_Position    0
Club_Joining     0
dtype: int64
```

Questions

Q1: What are people saying about me (somebody)?

We choose a person, and search hashtags ,tweets and retweets about her/him. Here we use "Lionel Messi" and "Cristiano Ronaldo" as two examples.

```
In [27]: db.soccer_twitter.find({'player_name':{'$regex':'Cristiano Ronaldo'}},{'id':1,'tweets':1,'_id':0,'retweet_count':1})
```

```
ernRealMadrid #BayernMunich #UCL #ChampionsLeaguÃ¢\x80A', 'retweet_count': 30}
{'id': 130, 'tweets': 'RT @FootballAdictt: Marco #Asenso Goal HD - Bayern Munich 1-2 Real Madrid 25.04.2018\n\n#Ba
ernRealMadrid #BayernMunich #UCL #ChampionsLeaguÃ¢\x80A', 'retweet_count': 8}
{'id': 98, 'tweets': 'Ã¢\x9fÃ¢\x92Ã¢ is coming....next season\n\n#ChampionsLeague #UCL #Cristiano #Ronaldo #Casillas
#FootTheBall https://t.co/vd4iqQf7aO', 'retweet_count': 2}
{'id': 134, 'tweets': 'RT @FootballAdictt: Marcelo Goal HD - Bayern Munich 1-1 Real Madrid 25.04.2018\n\n#Marcelo\n#Bay
ernRealMadrid #BayernMunich #UCL #ChampionsLeaguÃ¢\x80A', 'retweet_count': 30}
{'id': 155, 'tweets': 'RT @FootballAdictt: Marcelo Goal HD - Bayern Munich 1-1 Real Madrid 25.04.2018\n\n#Marcelo\n#Bay
ernRealMadrid #BayernMunich #UCL #ChampionsLeaguÃ¢\x80A', 'retweet_count': 30}
{'id': 96, 'tweets': 'Tampil Hebat, Rating Salah di FIFA 18 Melonjak Drastis!\n-----\n#Cristiano Ronaldo #FIFA 1
8 #Lionel MessiÃ¢\x80A; https://t.co/2wrEdC2I3Q', 'retweet_count': 0}
{'id': 85, 'tweets': 'who is shadow man of droba ? #messi #salah #ronaldo #barca #penalty #MessiReyBarÃ\x83Ã§a #messi
#penalty\n#christiano #lfcÃ¢\x80A; https://t.co/s86mu5CYeh', 'retweet_count': 1}
{'id': 70, 'tweets': 'RT @zonemixteFr: Ã¢\x9aÃ¢\x92Ã¢\x9a\x8f#ChampionsLeague\n\nÃ¢\x9fÃ¢\x92\x8fÃ¢\x86 Nouveau record pour #C
ristiano Ronaldo.\n\nGrÃ¢\x83Ã¢ce Ã¢\x83Ã¢\x80 la victoire du Real hier Ã¢\x83Ã¢\x80 Munich, le Portugais estÃ¢\x80A!', 
'retweet_count': 2}
{'id': 117, 'tweets': 'Cristiano Ronaldo has the most wins in the history of the European Cup\nÃ¢\x9fÃ¢\x94Ã¢ Footba
ll News: \n\n#wins #christianoÃ¢\x80A; https://t.co/Iw91ljlvVY', 'retweet_count': 0}
{'id': 66, 'tweets': 'RT @FootTheBall: Lucky Rashford! Ã¢\x9fÃ¢\x91Ã¢\x91Ã¢\x95\n\n#Rashford #Cristiano #Ronald #ManUtd #Re
alMadrid #FootTheBall https://t.co/OQxdFGb31A', 'retweet_count': 2}
```

```
In [29]: match = {
    'tweets': {"$regex": "Cristiano Ronaldo"}
}
```

```
group = {  
    '_id': "$tweets"  
}
```

```
ret2 = db.soccer_twitter.aggregate([
    {'$match':match},
    {'$group': group}
])
```

```
list(ret2)
```

Out[29]: [{ '_id': 'RT #nbayramberdiyev: #Cristiano Ronaldo's Stunning Bicycle Kick Goal Against Juventus 03.04.2018\n#UEFA Champions League #ChampionsLeague #UCL #A\x80!',
 { '_id': 'Superb sub Marco Asensio tops our #MadridistaReal player ratings as #Cristiano Ronaldo has a quiet night.\n\n\nhttps://t.co/ybb3sjLRDV',
 { '_id': 'Tampil Hebat, Rating Salah di FIFA 18 Melonjak Drastis!\n-----\n#Cristiano Ronaldo #FIFA 18 #Lionel Messi\n\x8c\x80; https://t.co/2wrEdC2I3Q',
 { '_id': 'Que Golazo hab\x83A\xada marcado Cristiano Ronaldo que se anula por apoyarse con la mano.\n\n#BayernMunich
1-2 #RealMadrid 70\x8c\x80; https://t.co/mPsSUG77I4',
 { '_id': 'RT @defcentral: El plan de Cristiano Ronaldo para llegar a la tope al choque de vuelta ante el Bayern: http://t.co/qTjk63NrvtN#realmadrid #CR\x8c\x80!',
 { '_id': 'Cristiano Ronaldo REACTION to Fans Which Entered The Stadium\x8c\x80; x9f\x8c\x80; x98\x8c\x80; x9f\x8c\x80; x98\x8c\x80;\nFull video : http://t.co/kbbgMd0TwN#APorLa13A\x8c\x80; https://t.co/3EZkdvjFii',
 { '_id': 'Les photos interdites de Georgina Rodriguez, la petite amie de Cristiano Ronaldo https://t.co/yJ3J55nqYes\n#Cristiano',
 { '_id': '#BabeKasihTau Sepatu Emas Eropa 2017-2018: Mohamed Salah dan Lionel Messi Terdepan, Cristiano Ronaldo Tak Mau asuk Hit\x8c\x80; https://t.co/OstNiC3sYG',
 { '_id': 'Cristiano Ronaldo has the most wins in the history of the European Cup\n\n#Football News:\n\n#wins #cristiano\x8c\x80; x80; https://t.co/Iw91ljlvvY',
 { '_id': '\x8c\x80; x9a\x8c\x80; \x8c\x80; \x8c\x80; x8f#ChampionsLeague\n\n#A\x8c\x80; x9f\x8c\x80; x8f\x8c\x80; x86 Nouveau record pour #Cristiano Ronaldo.\n\n#ngR\x8c\x80; x83

Q6:What users post like me?

We use regex to find similar users.

```
In [30]: ret2 = db.soccer_twitter.aggregate([
    {'$match': {'tweets': {'$regex': "Lionel Messi"}}},
    {'$group': {'_id': '$user_id'}}
])
list(ret2)
```

```
Out[30]: [ {_id: 202610992},
  {_id: 847142958},
  {_id: 717039627916484608},
  {_id: 231726084},
  {_id: 987538671288094720},
  {_id: 770751957245894656},
  {_id: 2306171660},
  {_id: 1355634392},
  {_id: 42480559},
  {_id: 4097611282},
  {_id: 834154403930238979},
  {_id: 776417140907749376},
  {_id: 929778826141454336},
  {_id: 815451895691087874}]
```

```
In [31]: ret2 = db.soccer_twitter.aggregate([
    {'$match': {'tweets': {'$regex': "Cristiano Ronaldo}}},
    {'$group': {'_id': '$user_id'}}
])
list(ret2)
```

```
Out[31]: [ {_id: 2306171660},
  {_id: 971400290795651072},
  {_id: 592426269},
  {_id: 4097611282},
  {_id: 2399322038},
  {_id: 804414323934175232},
  {_id: 144363080},
```

Q7.What topics are trending in my domain?

We can compare the change of hashtags among different time to predict future trending.

```
In [34]: list(db.soccer_twitter.aggregate([
    {"$sort": {"created_at": 1}}))
```

...

```
In [35]: for doc in collection.find({}, {"player_name": 1, "_id": 0, "created_at": 1}).sort('created_at', pymongo.ASCENDING):
    print(doc)
```

...

```
In [37]: for post in db.soccer_twitter.find({'player_name': {'$regex': 'Lionel Messi'}}, {'id': 1, 'player_name': 1, '_id': 0, 'created_at': 1}):
    print(post)
```

...

```
In [38]: list(collection.find({
    "created_at": {
        "$gte": "2018-04-16 14:04:07",
        "$lte": "2018-04-18 23:59:59"}},
    {"player_name": 1, "_id": 0, "created_at": 1}).sort('created_at', pymongo.ASCENDING))
```

...

```
In [39]: list(collection.find({
    "created_at": {
        "$gte": "2018-04-19 14:04:07",
        "$lte": "2018-04-21 23:59:59"}},
    {"player_name": 1, "_id": 0, "created_at": 1}).sort('created_at', pymongo.ASCENDING))
```

...

```
list(collection.find({
    "created_at": {
        "$gte": "2018-04-22 14:04:07",
        "$lte": "2018-04-24 23:59:59"}},
    {"player_name": 1, "_id": 0, "created_at": 1}).sort('created_at', pymongo.ASCENDING))
```

```
list(collection.find({
    "created_at": {
        "$gte": "2018-04-22 14:04:07",
        "$lte": "2018-04-24 23:59:59"}},
    {"player_name":1, "_id":0, "created_at":1}).sort('created_at', pymongo.ASCENDING))
```

```
In [41]: list(collection.find({
    "created_at": {
        "$gte": "2018-04-25 14:04:07",
        "$lte": "2018-04-26 23:59:59"}},
    {"player_name":1, "_id":0, "created_at":1}).sort('created_at', pymongo.ASCENDING))
```

...

Q8. What keywords/ hashtags should I add to my post?

By counting the sum of retweet count of the same tag and ordering it, we can know that the post contains which tags is popular in a specific domain. These tags are what we would like to add to our posts.

```
In [32]: list(db.soccer_twitter.aggregate([
    {"$group":{"_id":"$array_agg","count":{"$sum":1}}},
    {"$sort":{"count":-1}}]))
```

```
Out[32]: [{"_id": "[‘Marcelo’, ‘BayernRealMadrid’, ‘BayernMunich’, ‘UCL’]",
  "count": 24},
{"_id": "[‘spain’, ‘spanish’, ‘harryfan’, ‘harrycool15’, ‘realmadrid’, ‘laliga’, ‘garethbale11’, ‘bale’, ‘Wales’, ‘w
elsh’, ‘realmadridcf’, ‘gb11’]",
  "count": 15},
{"_id": “[‘Neymar’]”, “count”: 10},
{"_id": “[‘Asensio’, ‘BayernRealMadrid’, ‘BayernMunich’, ‘UCL’]”, “count”: 9},
{"_id": “[‘Neymar’, ‘PSG’]”, “count”: 6},
{"_id": “[‘psg’, ‘PSGASM’, ‘dimaria’, ‘cavani’, ‘Locelso’, ‘draxler’, ‘mbappe’, ‘Neymar’]”,
  “count”: 5},
{"_id": “[‘GARETH’, ‘BALE’, ‘PSICOLOGIA’]”, “count”: 5},
{"_id": “[‘Andres’, ‘Lionel’]”, “count”: 4},
{"_id": “[‘goals’, ‘messi’, ‘realmadrid’, ‘fcbarcelona’, ‘lionelmessi’, ‘leomessi’, ‘leo’]”,
  “count”: 4},
{"_id": “[‘ChampionsLeague’, ‘Cristiano’]”, “count”: 3},
{"_id": “[‘Salah’, ‘Neymar’, ‘Liverpool’]”, “count”: 3},
{"_id": “[‘Messi’, ‘Argentina’, ‘CA2016’]”, “count”: 3},
{"_id": “[‘APorLal3’]”, “count”: 3},
{"_id": “[‘Cristiano’]”, “count”: 3},
```

```
In [33]: list(db.soccer_twitter.aggregate([
    {"$match":{"player_name":“Lionel Messi”}},
    {"$group":{"_id":"$array_agg","count":{"$sum":1}}},
    {"$sort":{"count":-1}}]))
```

```
Out[33]: [{"_id": “[‘Andres’, ‘Lionel’]”, “count”: 4},
{"_id": “[‘lionel’]”, “count”: 3},
{"_id": “[‘Messi’, ‘Argentina’, ‘CA2016’]”, “count”: 3},
{"_id": “[‘Lionel’]”, “count”: 3},
{"_id": “[‘Goles’, ‘De’, ‘Messi’, ‘Desde’, ‘La’, ‘ESTADIO’, ‘FcBarcelona’, ‘Fcbarcelona’, ‘Grada’]”,
  “count”: 2},
{"_id": “[‘Madrid’, ‘Vs’, ‘Barcelona’, ‘Messi’, ‘Vs’, ‘Griezmann’, ‘All’]”,
  “count”: 2},
{"_id": “[‘FTPPro’, ‘LaLigaSantander’, ‘messi’, ‘lionelmessi’, ‘lionel’, ‘messi10’]”,
  “count”: 2},
{"_id": “[‘PES’, ‘Wii’, ‘Fc’, ‘Barcelona’, ‘PES’, ‘Champions’, ‘League’, ‘Andres’]”,
  “count”: 2},
{"_id": “[‘Lionel’, ‘Messi’, ‘vince’, ‘anche’]”, “count”: 2},
 {"_id": “[‘Fc’, ‘Barcelona’, ‘Abidal’, ‘Real’, ‘Madrid’, ‘Blaugrana’, ‘CF’, ‘CulÃ\x83Â©’, ‘Ã\x83\\\x89ric’, ‘Fifa’]”,
  “count”: 2},
 {"_id": “[‘traemelacopamessi’, ‘messi’, ‘mundial’, ‘russia2018’, ‘seleccion’, ‘argentina’, ‘lionel’, ‘cup’]”,
  “count”: 1},
 {"_id": “[‘Messi’, ‘Fcbarcelona’, ‘Estudiantes’, ‘BARCA’]”, “count”: 1},
 {"_id": “[‘Lionel’, ‘Messi’, ‘PaniniStickerAlbum’, ‘Rusia2018’]”, “count”: 1},
```